Integrating Automatic Differentiation with Object-Oriented Toolkits for High-Performance Scientific Computing

Jason Abate[1]
Steve Benson[2]
Lisa Grignon[3]
Paul Hovland[2]
Lois McInnes[2]
Boyana Norris[2]

ABSTRACT Often the most robust and efficient algorithms for the solution of large-scale problems involving nonlinear PDEs and optimization require the computation of derivative quantities. We examine the use of automatic differentiation (AD) to provide code for computing first and second derivatives in conjunction with two parallel numerical toolkits, the Portable, Extensible Toolkit for Scientific Computing (PETSc) and the Toolkit for Advanced Optimization (TAO). We discuss how the use of mathematical abstractions for vectors and matrices in these libraries facilitates the use of AD to automatically generate derivative codes and present performance data demonstrating the suitability of this approach.

1 Introduction

As the complexity of advanced computational science applications has increased, the use of object-oriented software methods for the development of both applications and numerical toolkits has also increased. The migration toward this approach can be attributed in part to encapsulation of data and algorithms and code reusability provided by well-designed ob-

[1]Texas Institute for Computational and Applied Mathematics, University of Texas, Austin, TX.
[2]Mathematics and Computer Science Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL 60439-4844.
[3]Mathematics Department, University of North Carolina, Chapel Hill, NC.

jects and object toolkits. Such toolkits enable developers to focus on a small component of a complex system, rather than attempting to develop and maintain a monolithic application. Furthermore, code reuse justifies expending significant effort in the development of highly optimized object toolkits encapsulating expert knowledge.

Many high-performance numerical toolkits include components designed to be combined with an application-specific nonlinear function. Examples include optimization components, nonlinear equation solvers, and differential algebraic equation solvers. Often the numerical methods implemented by these components also require first and possibly second derivatives of the function. Frequently, the toolkit is able to approximate these derivatives by using finite differences (FD); however, the convergence rate and robustness are often improved if the derivatives are computed analytically.

This represents an ideal situation for using automatic differentiation (AD) [Gri89, Gri00]. Developing correct parallel code for computing the derivatives of a complicated nonlinear function can be an onerous task, making an automated alternative quite attractive. Furthermore, the well-defined interfaces used by object toolkits simplify automatic differentiation by removing the need for the programmer to identify the independent and dependent variables and/or write code for the initialization of seed matrices.

We examine the use of automatic differentiation to provide code for computing first and second derivatives in conjunction with two numerical toolkits, the Portable, Extensible Toolkit for Scientific Computing (PETSc) and the Toolkit for Advanced Optimization (TAO). We describe how AD can be used with these toolkits to generate the code for computing the derivatives automatically. We present results demonstrating the suitability of AD and PETSc for the parallel solution of nonlinear PDEs and mention preliminary results from the use of AD with TAO.

2 Portable, Extensible Toolkit for Scientific Computing

PETSc [BGMS97, BGMS00] is a suite of data structures and routines for the scalable solution of scientific applications modeled by partial differential equations. The software integrates a hierarchy of components that range from low-level distributed data structures for vectors and matrices through high-level linear, nonlinear, and timestepping solvers. The algorithmic source code is written in high-level abstractions so that it can be easily understood and modified. This approach promotes code reuse and flexibility and, in many cases, helps to decouple issues of parallelism from algorithm choices.

Newton-based methods (see, e.g., [NW99]), which offer the advantage of rapid convergence when an iterate is near to a problem's solution, form the algorithmic core of the nonlinear solvers within PETSc. The methods employ line search, trust region, and pseudo-transient continuation strategies to extend the radius of convergence of the Newton techniques, and often solve the linearized systems inexactly with preconditioned Krylov methods. The basic Newton method requires the Jacobian matrix, $J = F'(u)$, of a nonlinear function $F(u)$. Matrix-free Newton-Krylov methods require Jacobian-vector products, $F'(u)v$, and may require an approximate Jacobian for preconditioning.

3 Toolkit for Advanced Optimization

TAO [BMM99, BMM00] focuses on scalable optimization software, including nonlinear least squares, unconstrained minimization, bound constrained optimization, and general nonlinear optimization. The TAO optimization algorithms use high-level abstractions for matrices and vectors and emphasize the reuse of external tools where appropriate, including support for using the linear algebra components provided by PETSc and related tools.

Many of the algorithms employed by TAO require first and sometimes second derivatives. For example, unconstrained minimization solvers that require the gradient, $f'(u)$, of an objective function, $f(u)$, include a limited-memory variable metric method and a conjugate gradient method, while solvers that require both the gradient, $f'(u)$, and Hessian, $f''(u)$, (or Hessian-vector products) include line search and trust region variants of Newton methods. In addition, algorithms for nonlinear least squares and constrained optimization often require the Jacobian of the constraint functions or at least the computation of Jacobian-vector and Jacobian-transpose-vector products.

4 Using Automatic Differentiation

The automatic differentiation tools used in this research are the ADIFOR (Automatic Differentiation of Fortran) [BCKM96] and ADIC (AD of C) [BRM97] systems. Given code for a function $f(u)$ in Fortran 77 or ANSI C, these tools generate code for the computation of $f'(u)$ and, if desired, $f''(u)$.

5 Experimental Results

Other work [LP99, FMM98, Ger00, LH00] has demonstrated the benefits of well-defined interfaces for automating the AD process. The object-oriented

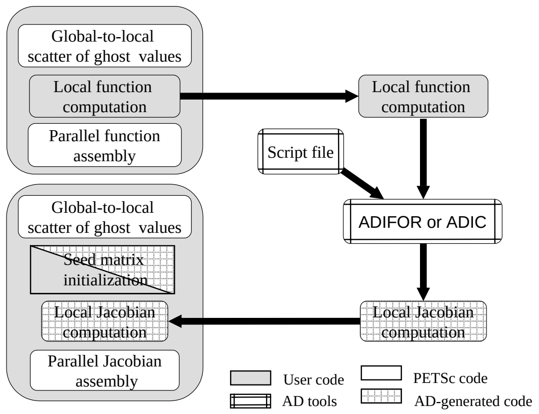
Global-to-local scatter of ghost values
Local function computation
Parallel function assembly
Local function computation
Script file
ADIFOR or ADIC
Global-to-local scatter of ghost values
Seed matrix initialization
Local Jacobian computation
Parallel Jacobian assembly
Local Jacobian computation
User code
PETSc code
AD tools
AD-generated code

FIGURE 1. Schematic diagram of the use of automatic differentiation tools to generate the Jacobian routine for a nonlinear PDE computation.

designs of PETSc and TAO lead to such well-specified interfaces. However, rather than exploit this feature directly, we have chosen to take advantage of the structure of a typical PETSc/TAO nonlinear function evaluation to simplify the AD process. The parallel nonlinear function code usually includes several calls to PETSc/TAO routines for generalized vector scatters before and after the actual local function computation. These calls take care of data structure setup and communication, enabling a completely local function computation. In the current semi-automatic approach, illustrated in Figure 1, we differentiate this local function using AD. This produces code for local derivative computation, which is coupled with code for assembling the gradient, Jacobian, or Hessian. While in principle it is possible to differentiate through the parallel scatter and assembly routines [Car, FD99, Hov97, HB98], currently the corresponding seed-matrix initialization and assembly code are generated manually. Future development will automate this process.

We present experimental results for nonlinear PDEs and unconstrained minimization problems that demonstrate the utility of AD in conjunction with parallel numerical libraries. These computations were run on an IBM SP with 120 MHz P2SC nodes with two 128 MB memory cards each and a TB3 switch. We have observed analogous qualitative behavior on a range of other current parallel architectures.

5.1 Using AD and PETSc for Nonlinear PDEs

We used AD and PETSc to solve the steady-state, three-dimensional compressible Euler equations on mapped, structured meshes using a second-

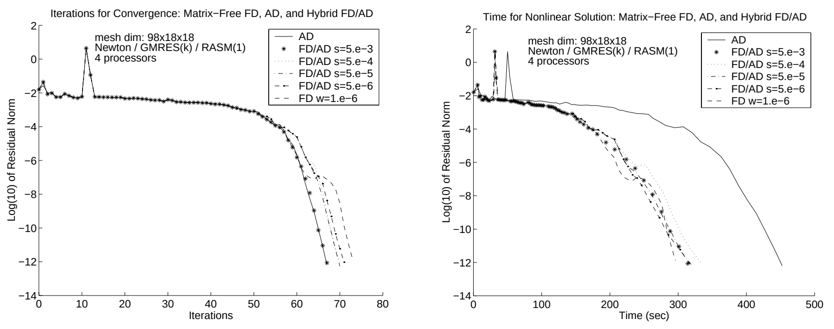

FIGURE 2. Comparison of convergence using matrix-free Newton-Krylov--Schwarz methods with finite differencing (FD), automatic differentiation (AD), and hybrid variants (FD/AD) with various switching parameters, s. Iterations (left) and time (right).

order, Roe-type, finite-volume discretization. In particular, we solved in parallel a system of the form $F(u) = 0$, where $F : \Re^n \rightarrow \Re^n$, using matrix-free Newton-Krylov-Schwarz algorithms with pseudo-transient continuation to model transonic flow over an ONERA M6 airplane wing. See [GKMT98] for details about the problem formulation and algorithmic approach. The linearized Newton correction equations were solved by using restarted GMRES preconditioned with the restricted additive Schwarz method with one degree of overlap.

As discussed in depth in [HM00] and summarized in Figure 2, our results indicate that, within the context of matrix-free Newton-Krylov methods, AD offers significantly greater robustness and converges in fewer iterations than FD. This figure plots convergence rate in terms of the residual norm $||F(u)||_2$ versus both nonlinear iteration number and computation time on four processors for a model problem of dimension 158,760. The runtime for AD is slightly higher than for FD (when an appropriate differencing stepsize is used), due to the higher cost of computing Jacobian-vector products using AD. However, coupling AD with FD in a hybrid scheme provides the robustness of AD with the lower computation time of FD, without needing to identify the proper stepsize for FD. Additional experiments show that this hybrid technique scales well for various problem sizes and processor configurations [HM00].

5.2 *Using AD and TAO for Unconstrained Minimization*

We evaluated the preliminary performance of automatic differentiation in conjunction with TAO using a two-dimensional elastic-plastic torsion model from the MINPACK-2 test problem collection [ACMX92]. This model uses a finite element discretization to compute the stress field on an infinitely long cylindrical bar to which a fixed angle of twist per unit length has

TABLE 1.1. Execution times (sec) for various stages of the elastic-plastic torsion minimization problem.

Number of Processors	1		16		32	
Number of Vertices	10,000		160,000		320,000	
Hessian Computation Method	FD	AD	FD	AD	FD	AD
Linear System Solution	1.19	1.20	7.10	7.04	9.16	9.04
Compute Function	0.01	0.01	0.04	0.02	0.04	0.02
Compute Gradient	0.21	0.20	0.22	0.24	0.23	0.24
Compute Hessian	1.89	1.48	3.86	1.58	6.11	1.65
Total Time	5.20	4.86	14.13	11.87	18.73	14.32

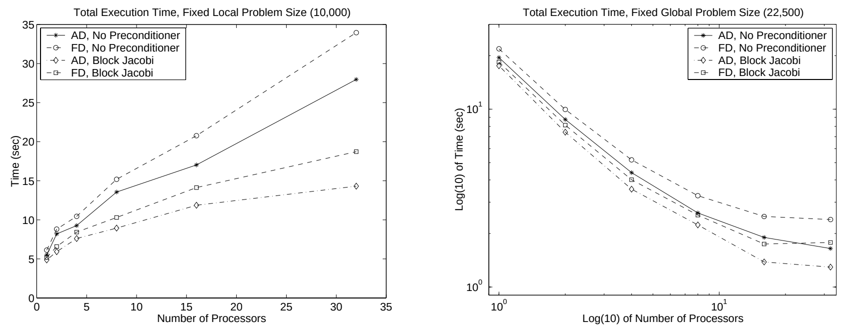

FIGURE 3. Comparison of total execution times for the elastic-plastic torsion minimization problem using a line search Newton method. Fixed local (left) and fixed global (right) problem sizes.

been applied. The resulting unconstrained minimization problem can be expressed as $min\ f(u)$, where $f : \Re^n - > \Re$. All of the following numerical experiments use a line search Newton method with a preconditioned conjugate gradient linear solver.

We compared two approaches for computing the full Hessian of $f(u)$. First, we applied AD to the hand-coded routine computing the analytic gradient of $f(u)$. Second, we used FD to approximate the Hessian, again using the analytic gradient of $f(u)$. In both cases we employed graph coloring techniques in order to exploit the sparsity of the Hessian computation [CM83, GT84]. The graphs in Figure 3 show the scaling of the complete minimization problem using either no preconditioning or a block Jacobi preconditioner. The block Jacobi preconditioner uses a subdomain solver of ILU(0) on each block, with one block per processor. For this relatively simple problem, both AD and FD exhibit rapid convergence in terms of number of iterations. However, AD outperforms FD in terms of total time to solution, mainly because of the good scalability of the AD Hessian computation. Overall, the results for a fixed local 100×100 mesh size indicate that the problem does not scale well for either AD or FD (although using a block Jacobi preconditioner helps somewhat). This situation is due in part to the poor performance scaling of the linear system solution (see Table 1.1). We are currently exploring the causes of this poor scalability.

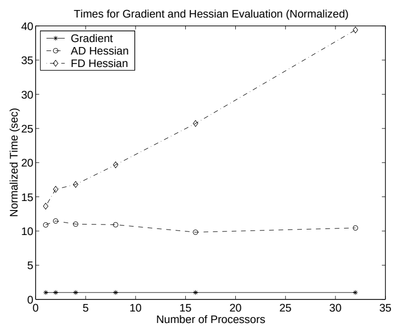

FIGURE 4. Comparison of sparse Hessian evaluation times for a fixed local problem size (10,000) using a line search Newton method and a block Jacobi preconditioner with FD and AD.

Figure 4 contains the time for a single Hessian evaluation, normalized by the time for a single gradient evaluation. The AD Hessian computation scales very well for the problem and machine sizes in our experiments, while FD with coloring fails to achieve good parallel performance. We attribute the poor scalability of FD to the collective operations used in the FD step-size determination. Even on one processor, AD offers superior performance; this could be due to better cache performance because of increased data and temporal locality. Further investigations are necessary to fully understand the sources of the AD performance advantage.

6 Summary and Future Work

We have presented a methodology for using AD to compute first and second derivatives for use in the parallel solution of nonlinear PDEs and optimization. The robustness and, in some cases, performance of the resulting code are superior to results with finite difference approximations. Also, in contrast to hand-coding, AD can easily be re-applied if the nonlinear function changes.

Future work includes moving from a semi-automatic approach to an approach, in which code for seed matrix initialization and derivative matrix assembly are completely automated. This task will be aided by the existence of well-defined interfaces for the nonlinear function component and will leverage work on developing a differentiated version of PETSc [HNRS98].

Acknowledgments

This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

We thank Jorge Moré and Barry Smith of Argonne National Laboratory for stimulating discussions about issues in derivative computations with PDE and optimization software and Gail Pieper for proofreading an early draft of this manuscript.

7 References

[ACMX92] B. M. Averick, R. G. Carter, Jorge J. Moré, and G.-L. Xue. The MINPACK-2 test problem collection. Preprint MCS-P153-0694, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, 1992.

[BCKM96] Christian Bischof, Alan Carle, Peyvand Khademi, and Andrew Mauer. ADIFOR 2.0: Automatic differentiation of Fortran 77 programs. *IEEE Computational Science & Engineering*, 3(3):18–32, 1996.

[BGMS97] Satish Balay, William D. Gropp, Lois Curfman McInnes, and Barry F. Smith. Efficient management of parallelism in object oriented numerical software libraries. In E. Arge, A. M. Bruaset, and H. P. Langtangen, editors, *Modern Software Tools in Scientific Computing*, pages 163–202. Birkhauser Press, 1997.

[BGMS00] Satish Balay, William D. Gropp, Lois Curfman McInnes, and Barry F. Smith. PETSc 2.0 users manual. Technical Report ANL-95/11 - Revision 2.0.28, Argonne National Laboratory, March 2000. See `http://www.mcs.anl.gov/petsc`.

[BMM99] Steve Benson, Lois Curfman McInnes, and Jorge Moré. GPCG: A case study in the performance and scalability of optimization algorithms. Technical Report ANL/MCS-P768-0799, Mathematics and Computer Science Division, Argonne National Laboratory, 1999.

[BMM00] Steve Benson, Lois Curfman McInnes, and Jorge Moré. TAO users manual. Technical Report ANL/MCS-TM-242, Mathematics and Computer Science Division, Argonne National Laboratory, 2000. See `http://www.mcs.anl.gov/tao`.

[BRM97] Christian Bischof, Lucas Roh, and Andrew Mauer. ADIC — An extensible automatic differentiation tool for ANSI-C. *Software–Practice and Experience*, 27(12):1427–1456, 1997.

[Car] Alan Carle. Personal communication.

[CM83] T. F. Coleman and J. J. Moré. Estimation of sparse Jacobian matrices and graph coloring problems. *SIAM J. Numer. Anal.*, 20:187–209, 1983.

[FD99] Christèle Faure and Patrick Dutto. Extension of Odyssée to the MPI library - direct mode. Technical Report 3715, Institut National de Reserche en Informatique et en Automatique, 1999.

[FMM98] M. C. Ferris, M. Mesnier, and J. J. Moré. NEOS and Condor: Solving optimization problems over the Internet. Preprint ANL/MCS-P708-0398, MCS Division, Argonne National Laboratory, Argonne, Ill., 1998.

[Ger00] Michael Gertz, 2000. Personal communication.

[GKMT98] W. D. Gropp, D. E. Keyes, L. C. McInnes, and M. D. Tidriri. Globalized Newton-Krylov-Schwarz algorithms and software for parallel implicit CFD. Technical Report 98-24, ICASE, August 1998. To appear in *Int. Journal on Supercomputing Applications*.

[Gri89] Andreas Griewank. On automatic differentiation. In *Mathematical Programming: Recent Developments and Applications*, pages 83–108, Amsterdam, 1989. Kluwer Academic Publishers.

[Gri00] Andreas Griewank. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia, 2000.

[GT84] D. Goldfarb and Ph. L. Toint. Optimal estimation of Jacobian and Hessian matrices that arise in finite difference calculations. *Mathematics of Computation*, 43:69–88, 1984.

[HB98] Paul Hovland and Christian Bischof. Automatic differentiation of message-passing parallel programs. In *Proceedings of the First Merged International Parallel Processing Symposium and Symposium on Parallel and Distributed Processing*, Los Alamitos, CA, 1998. IEEE Computer Society Press.

[HM00] P. D. Hovland and L. C. McInnes. Parallel simulation of compressible flow using automatic differentiation and PETSc. Technical Report ANL/MCS-P796-0200, Mathematics and Computer Science Division, Argonne National Laboratory, 2000. To appear in a special issue of *Parallel Computing* on "Parallel Computing in Aerospace".

[HNRS98] Paul Hovland, Boyana Norris, Lucas Roh, and Barry Smith. Developing a derivative-enhanced object-oriented toolkit for scientific computations. In *Proceedings of the SIAM Workshop on Object Oriented Methods for Inter-operable Scientific and Engineering Computing*, pages 129–137. SIAM, Oct 1998.

[Hov97] Paul D. Hovland. *Automatic Differentiation of Parallel Programs.* PhD thesis, University of Illinois at Urbana-Champaign, Urbana, IL, May 1997.

[LH00] Steven L. Lee and Paul D. Hovland. Sensitivity analysis using parallel ODE solvers and automatic differentiation in C: SensPVODE and ADIC. Technical Report ANL/MCS-P818-0500, Mathematics and Computer Science Division, Argonne National Laboratory, 2000. Submitted to AD2000.

[LP99] Shengtai Li and Linda Petzold. Design of new DASPK for sensitivity analysis. Technical report, University of California at Santa Barbara, 1999.

[NW99] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization.* Springer-Verlag, New York, 1999.